



Validating end of text packages using open source

Katrine Facius
Global Biometrics
Ferring Pharmaceuticals

Helping people live better lives



Outline

Validating end of text packages using open source

What is the problem?

- End Of Text (EOT) packages – a key deliverable for statistical programmers
- The validation and continuous updating of EOT packages

What is our solution?

- Getting the right input to get the right solution
- Exploring the world of open source tools
- The first implementation

What lies ahead?

- User friendliness
- Additional features – machine learning, AI, ...?

What is the problem?

End Of Text packages – a key deliverable for statistical programmers

- The EOT packages is one of the core trial deliveries for statistical programmers
- Data collected during a trial are to be part of the final clinical trial report
 - The accumulated information is presented through tables, listings and figures as addendum
 - The output document can easily be > 1000 pages long
- These packages needs to contain all the
 - necessary information
 - correct information
 - consistent information



What is the problem?

The validation and continuous updating of EOT packages

- The validation of these text packages are of great importance
 - A manual process with many and very time-consuming reviews
 - Validation by programmers, statisticians, medical writers, medical monitors and other stakeholders



Images by gpoinstudio on Freepik

What is the problem?

The validation and continuous updating of EOT packages

- The validation of these text packages are of great importance
 - A manual process with many and very time-consuming reviews
 - Validation by programmers, statisticians, medical writers, medical monitors and other stakeholders
- Validation today
 - Finding the needle in the haystack!
 - Updates or changes to titles, labels, etc.
 - Going through several validation rounds requires a tracking of the updates or changes to tables

Body measurements by age stratum and visit - FAS

Age stratum: All

	Placebo (N=107)	Trt 1 (N=111)	Trt 2 (N=103)	Trt 3 (N=97)	Total (N=418)
Screening					
Height (m)					
n	108	111	103	97	418
Mean (SD)	1.65 (0.07)	1.65 (0.07)	1.66 (0.07)	1.67 (0.07)	1.65 (0.07)
Median	1.65	1.65	1.66	1.67	1.66
Weight (kg)					
n	107	111	103	97	418
Mean (SD)	63.7 (8.9)	65.7 (10.4)	64.3 (9.6)	65.8 (12.0)	65.3 (10.7)
Median	62.6	63.8	63.0	62.6	63.0

N = Number of subjects

n = Number of subjects with observation

FEXXXXXX/YYYYYY/Report/01SEP2022-PROG/T_ADBL.sas

What is our solution?

Getting the right input to get the right solution

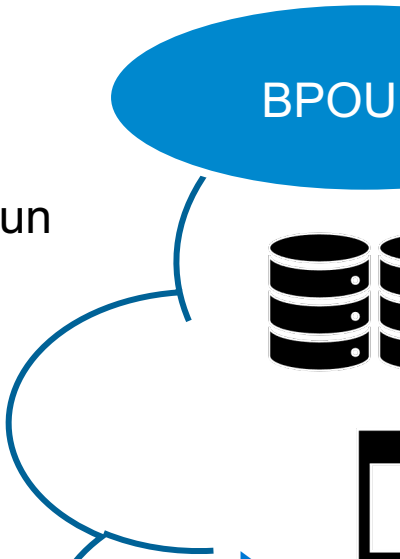
- Developing the right solution depends on getting the right input
 - Interview with statistical programmers to get input to problems that we could tackle
 - Instead we got a request to make a tool to replace the current Word compare!
- Need tool that can
 - pinpoint changes from version to version of these packages 
 - deliver an overview of changes to numbering, titles, lables, etc for the medical writer
- You need to listen to the end users!
 - Develop first version to replace existing tool (Word compare)
 - Once programmers are using this, we can further expand

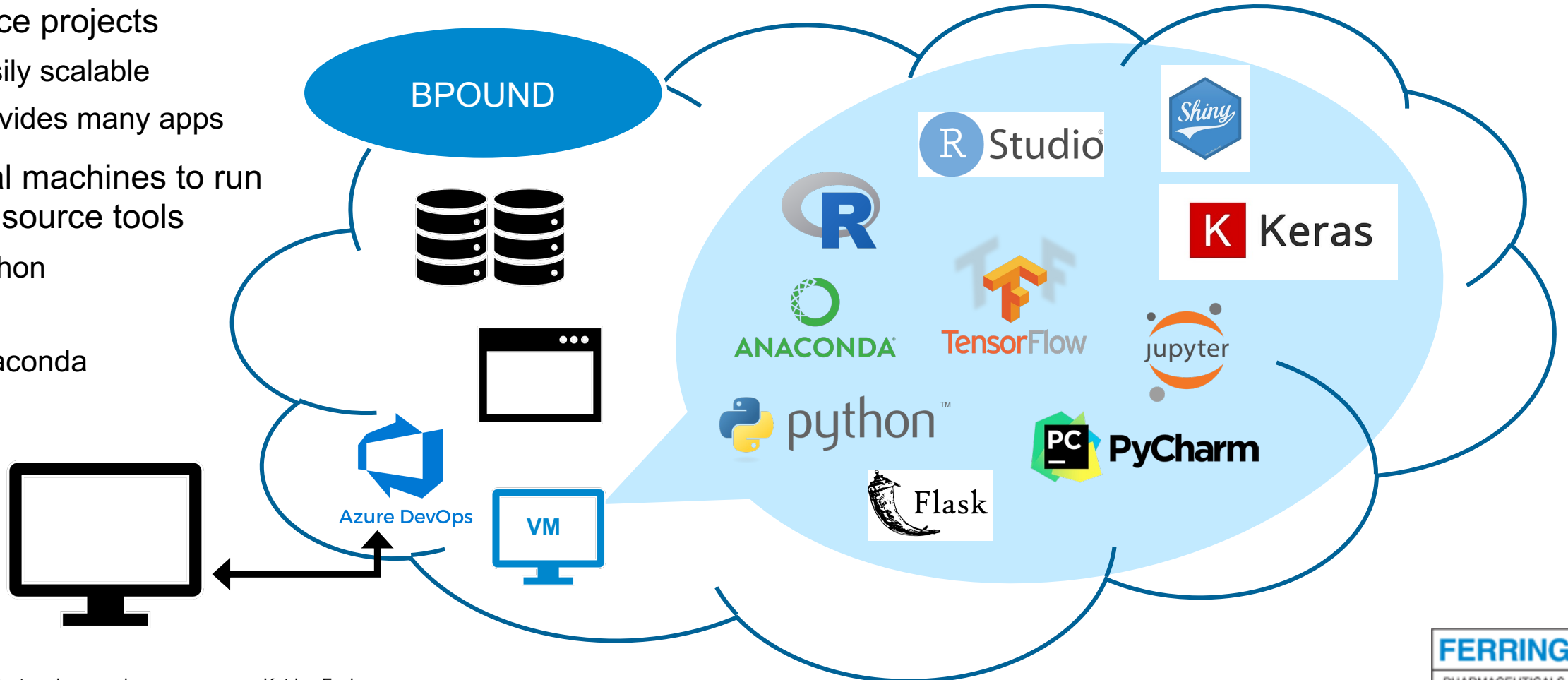


Designed by pch.vector / Freepik

What is our solution?

Exploring the world of open source tools

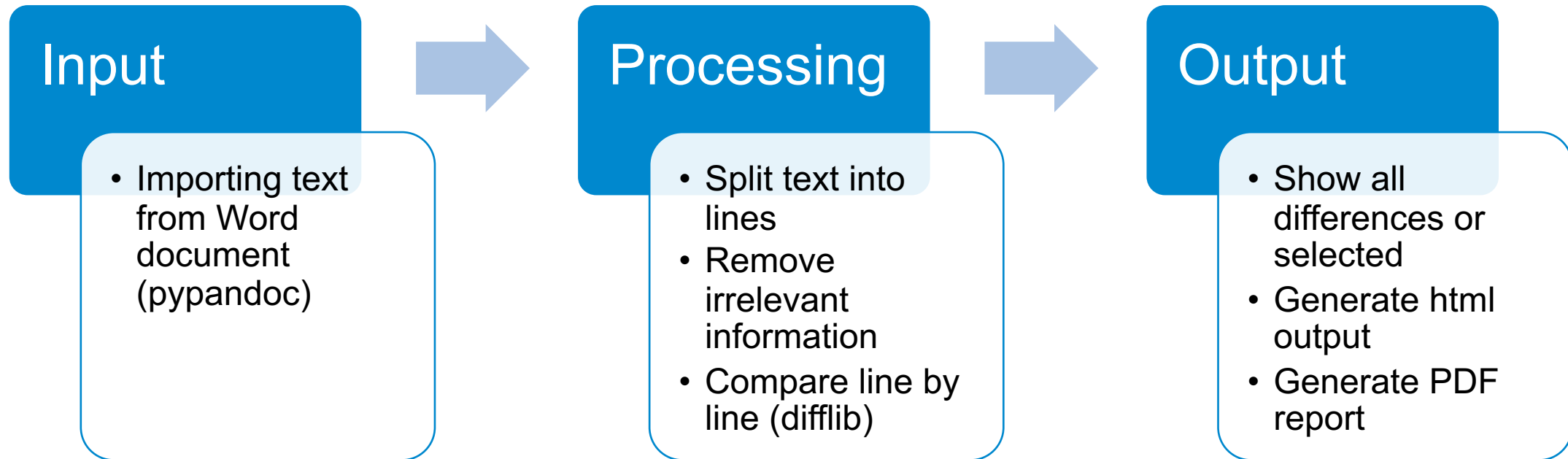
- BPOUND is our Azure platform for data science projects
 - Easily scalable
 - Provides many apps
 - Virtual machines to run open source tools
 - Python
 - R
 - Anaconda
 - ...
- 
- The diagram illustrates the BPOUND platform architecture. A blue oval at the top right is labeled 'BPOUND'. Below it, there are two database icons (cylinders) and a laptop icon. A blue line connects the 'BPOUND' oval to the database icons, and another blue line connects the database icons to the laptop icon.



What is our solution?

The first implementation

- Using Python and available libraries



What is our solution?

The first implementation

➤ Example of HTML output

1 Body measurements by age stratum and visit - FAS

2

3 Age stratum: All

4

5

6

7

8

9

10

11 Screening

12

13 Height (m)

14 n 108 111 103 97 418

15 Mean (SD) 1.65 (0.07) 1.65 (0.07) 1.66 (0.07) 1.67 (0.07) 1.65 (0.07)

16 Median 1.65 1.65 1.66 1.67 1.66

17

18 Weight (kg)

19 n 107 111 103 97 418

20 Mean (SD) 63.7 (8.9) 65.7 (10.4) 64.3 (9.6) 65.8 (12.0) 65.3 (10.7)

21 Median 62.6 63.8 63.0 62.6 63.0

22

23

24 N = Number of subjects

25 n = Number of subjects with observation

26

27 PROG/T_ADBL.sas

Legends

Colors	Links
Added	(f) first change
Changed	(n) next change
Deleted	(t) top

1 Body measurements by age stratum and visit - Full Analysis Set

2

3 Age stratum: Both

4

5

6

7

8

9

10

11 Screening visit

12

13 Height (m)

14 n 107 111 103 97 418

15 Mean (SD) 1.65 (0.07) 1.65 (0.07) 1.66 (0.07) 1.67 (0.07) 1.65 (0.07)

16 Min; max 1.5; 1.9 1.5; 1.8 1.5; 1.8 1.5; 1.8 1.5; 1.9

17

18 Weight (kg)

19 n 107 111 103 97 418

20 Mean (SD) 63.7 (8.9) 65.7 (10.4) 64.3 (9.6) 65.8 (12.0) 65.3 (10.7)

21 Min; max 44; 90 47; 93 46; 99 43; 100 43; 110

22

23

24 N = Number of subjects

25 n = Number of subjects with observation

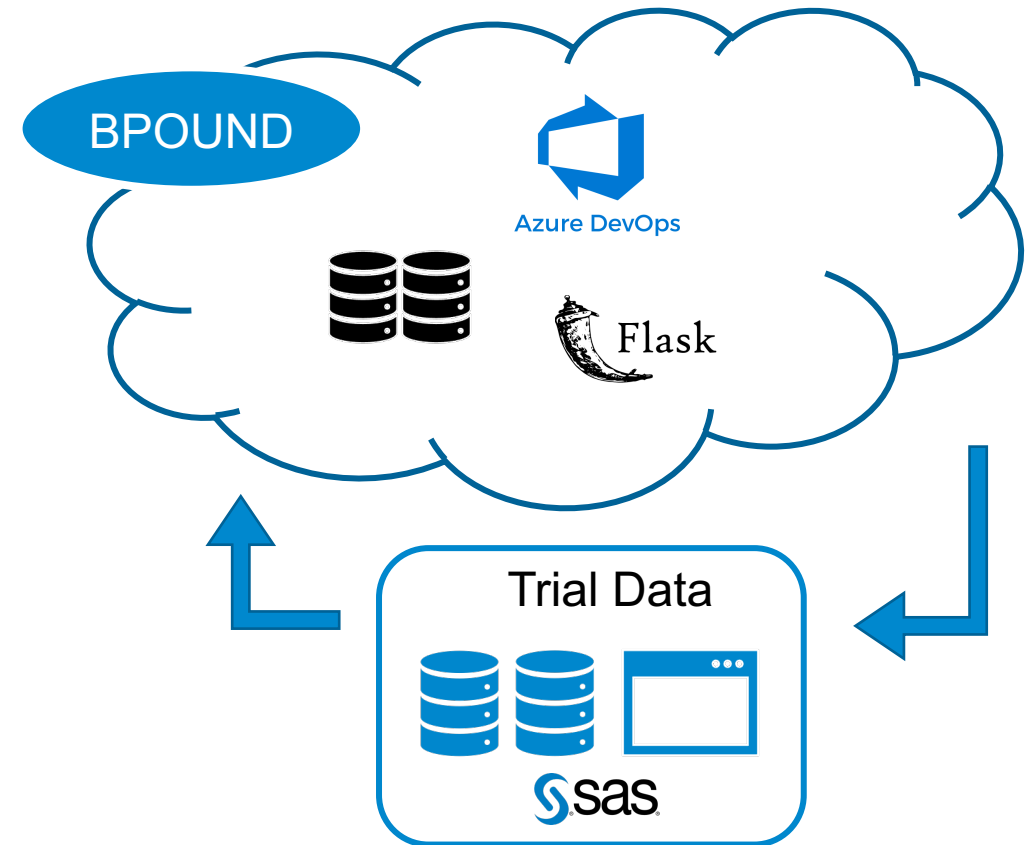
26

27 PROG/T_ADBL.sas

What lies ahead?

User friendliness

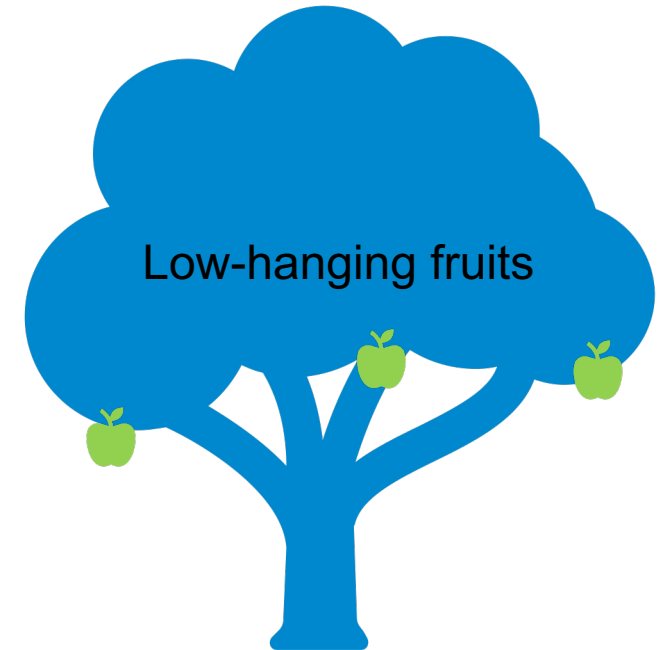
- The tool needs to work in a well integrated way with the systems the programmers usually use
- Getting data from our trial data system
 - Safety first!
 - Communicating directly from the trial system to the tool
- Web interface
 - Queuing and monitoring jobs
 - Presenting the result on a homepage
- Output
 - Delivering html based output
 - Delivering reports (PDF) from BPOUND and back to LSAF
- Validation of validation tool...



What lies ahead?

Additional features – logical checks

- Once we have a working tool that is preferred by the programmers we can add features
- Input from programmers on logical checks
 - Tables only containing zeros
 - Populations agree across tables
 - Logics based on therapeutic knowledge
- Can (easily) be added to the Python code since we already have the text strings



What lies ahead?

Additional features – machine learning, AI, ...

- Collect real examples of deviations
- Describe patterns we want to learn and deviations we want to be able to identify
- Design input representation and select initial model type
- Develop ML model
- Evaluate output
- Many models
 - NLP; Bag of Words
 - RNN
 - ...

doc_id 456



Learnings

Gains and obstacles in using open source in a pharma setup

- Expanding and exploring new ways to handle data in pharma
 - Better core deliveries
 - New possibilities
- Using open source has both upsides and downsides
 - A great and expanding portfolio of ready to use libraries, algorithms, machine learning models, apps and (relatively) easily implementable solutions
 - Interfacing with existing systems pose some difficulties
- We need to explore and expand the use of open source!



Designed by vectorjuice / Freepik

Connect with us!



katrine.facius@ferring.com



Ferring Pharmaceuticals



[@ferring](https://twitter.com/ferring)



Ferring Pharmaceuticals



[@ferringpharmaceuticals](https://www.instagram.com/ferringpharmaceuticals)



Ferring Pharmaceuticals

